hardware system that can fully support the the planned system and which can be integrated into a Hospital Information System. For this purpose a VAX 750 and three Xerox 1186 workstations have been acquired and our development efforts have been transferred to them.

D. Recommendations for Future Community and Resource Development

In the time we have been associated with SUMEX, we have been generally pleased with the facilities and services. However, it is clearly evident that the users' almost insatiable demands for CPU cycles and disk space cannot be met by a single central machine. The best strategy would appear to be one of emphasizing powerful workstations or relatively small, multi-user machines linked together in a nationwide network with SUMEX serving as the its central hub. This would give the individual users much more control over the resources available for their needs, yet at the same time allow for the communications among users that have been one of SUMEX's strong points.

For such a network to be successful, further work needs to be done in improving the network capabilities of SUMEX to encourage users at sites other than Stanford. Further work is also needed in the area of personal workstations to link them to such a network.

IV.C. Pilot Stanford Projects

Following are descriptions of the informal pilot projects currently using the Stanford portion of the SUMEX-AIM resource, pending funding, full review, and authorization.

IV.C.1. REFEREE Project

Principal Investigator: Bruce G. Buchanan, Ph.D.
Computer Science Department
Stanford University

Co-Principal Investigator: Byron W. Brown, Ph.D.
Department of Medicine
Stanford University

Associate Investigator: Daniel E. Feldman, Ph.D., M.D.
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

The goals of this project are related both to medical science and artificial intelligence: (a) use AI methods to allow the informed but non-expert reader of the medical literature to evaluate a randomized clinical trial, and (b) use the interpretation of the medical literature as a test problem for studies of knowledge acquisition and fusion of information from disparate sources. REFEREE and REVIEWER, a planned extension, will be used to evaluate the medical literature of clinical trials to determine the quality of a clinical trial, make judgments on the efficacy of the treatment proposed, and synthesize rules of clinical practice. The research is an initial step toward a more general goal - building computer systems to help the clinician and medical scientist read the medical literature more critically and more rapidly for use in making clinical decisions.

B. Medical Relevance

The explosive growth of the medical literature has created a severe information gap for the busy clinician. Most physicians can afford neither the time required to study all the pertinent journal articles in their field, nor the risk of ignoring potentially significant discoveries. The majority of clinicians, in fact, have little sophistication in epidemiology and statistics; they must nonetheless base their pragmatic decisions on a combination of clinical experience and published literature. The clinician's computerized assistant must ferret out useful maxims of clinical practice from the medical literature, pass judgment on the quality of medical reports, evaluate the efficacy of proposed treatments, and adjudicate the interpretation of conflicting and even contradictory studies.

C. Highlights of Progress

REFEREE presently encodes the methodological knowledge of a highly regarded biostatistician at Stanford (Dr. Bill Brown). The system allows the informed but non-expert reader of the medical literature to evaluate the credibility of a randomized clinical trial.

In the future, REFEREE and its extensions will alleviate the knowledge-acquisition bottleneck for an automated medical decision-maker: the program will help a reader to evaluate the quality of a clinical trial, judge the efficacy of the treatment proposed therein, and synthesize rules of clinical practice. For the present, however, the fusion of knowledge from disparate sources remains a problem in pure AI. The current effort of the REFEREE team is the appropriate representation of biostatistical knowledge in order to accomplish this set of tasks.

The REFEREE prototype is a consultant that evaluates the design and reporting of a single conclusion from randomized control trial for its credibility. It contains, in preliminary form, Professor Brown's expert knowledge of biostatistics. Given the assessments of the reader of various details, REFEREE synthesizes those judgments into a measure of credibility of the entire study. The reader may change his assessments in accordance with his uncertainty as to the judgments, and view graphically the resulting changes in REFEREE's measure.

The Knowledge Base:

Randomized controlled trials are used to test hypotheses regarding the effectiveness of various kinds of medical interventions. Dr. Brown classifies studies on the basis of three major attributes: the type of intervention tested (e.g., drug, surgery, health process change, etc.); the type of endpoint against which that intervention was tested (e.g. mortality, objective morbidity, subjective morbidity, etc.); and the type of conclusion drawn by the investigator/author on the basis of the research (e.g., that different treatments do or do not produce different outcomes, that a particular treatment is or is not cost-effective, etc.). Following this classificatory scheme, we decided to begin by producing a prototype REFEREE system that would help the reader to evaluate a single published conclusion concerning the effect of a given drug treatment on mortality.

Knowledge Acquisition:

Having defined the scope of the initial knowledge base, we turned to the problem of collecting the information from Dr. Brown for inclusion in the system, i.e., knowledge acquisition. This task generally involves a relatively long-term process of face-to-face information gathering during sessions between the expert and one or more knowledge engineers. Dr. Diana Forsythe has noted a parallel between the communicative and analytical tasks involved in knowledge acquisition and those undertaken in ethnographic research. For this reason, we included an anthropologist in the research team and make use of ethnographic techniques in order to maximize the efficiency and quality of the data collection process.

Dr. Lehmann and Dr. Forsythe have carried out a year of systematic interviews with Dr. Brown in order to begin the process of constructing and refining the knowledge base for the current REFEREE prototype. We have combined a case-based approach that allows us actively to observe Dr. Brown as he reads papers, with semi-directed interviewing oriented toward understanding his terminology and category system. We find that these techniques work very well: Dr. Brown's interest in the knowledge acquisition process has been sustained, and indeed has increased over time as the system based on his expertise has evolved. He is clearly comfortable with this approach, and notes that it has actually afforded him additional insight into the way he interprets the literature.

Over the course of the project, we have altered our knowledge representation from that of rules to that of an *influence diagram*. This is an acyclic directed graph of propositions or variables connected by links, where the absence of a link indicates conditional independence of the two variables. This formalism has been used in decision analysis to enable experts to convey their knowledge of a domain, and, more recently, has been used in AI to represent that knowledge in expert systems. This shift in formalism significantly altered the knowledge acquisition process and the implementation of that knowledge in our program.

Based on information from our expert, we have taken *credibility* as the goal parameter of the present system. This goal is defined operationally by Dr. Brown as "my odds that the conclusion of interest would be replicated in an experiment based on the methods reported in the paper but without any of the flaws". In assessing *credibility*, for instance, Dr. Brown considers the blindedness of the randomization, the blindedness of the execution, the equivalence of the two groups at baseline, the equivalence in treatment of the two groups, the completeness of results reporting, and the propriety of the statistical analysis. We recognize that these variables are not all conditionally independent on *credibility*; work is in progress to assess as accurately as possible just what the conditional relationships are. Our use of influence diagrams has numerous advantages: the approach is acceptable to Dr. Brown, it is flexible, it can represent several aspects of the structure of the knowledge used by the expert, and the resultant data can be entered easily into the computer.

Inference in REFEREE:

REFEREE was originally built within EMYCIN, a backward-chaining rule-based AI environment developed from MYCIN at Stanford. This environment is ideally suited for ordered collection of evidence and a diagnosis of the goal state at the end of that process. The state of belief or knowledge in parameters not directly between the evidence and the goal state is irrelevant. Our present system focuses on maintaining consistency over the entire knowledge base as new evidence is incorporated into the system. The constraints implied by the new data and Dr. Brown's prior knowledge are propagated throughout the system by Judea Pearl's message-passing algorithm for belief networks. During the consultation with the program, questions are chosen by the user and answered at his or her discretion, and the state of belief in any parameter can be requested at any time. The odds of replicating the study, then, can be viewed at any point during evidence collection.

There are a number of choices in representing our domain in an influence diagram. One is to view the goal of *credibility* as a proposition, the uncertainty in which is calculated by Pearl's algorithm. In this case, there are two choices: to view important design and execution factors as conditionally independent, given an assessment of the *credibility*, or to view them as causal of the goal measure. The program is currently implemented in the first topology, and we plan to test the second as well. A second representation is to view *credibility* as a measure of value, in which case the current knowledge base represents an objectives hierarchy, in the language of multi-attribute decision theory. We implemented REFEREE in David Klein's VERTUS system, following this paradigm, with moderate improvement in REFEREE's explanatory power.

A third representation is to reinterpret the task of REFEREE entirely, and to view it in the context of a physician's decision to treat or not to treat a patient with the intervention tested in the study under consideration. Dr. Lehmann is exploring this interpretation for his doctoral thesis.

The User Interface:

REFEREE was initially run entirely on the SUMEX resource. Mr. Chavez reimplemented the program on a stand-alone workstation, the Xerox 1186 in the KEE commercial expert system shell. The availability of bit-mapped screens made us more sensitive to issues of the user interface, but the shell could not deal easily with the uncertainty inherent in our domain. Mr. Chavez then ported the system to a Texas Instrument Explorer work-station, for which he designed an entirely new knowledge engineering shell which integrated EMYCIN and influence diagrams. It was apparent, however, that to accommodate the multiple interface needs of our potential user community, we needed a graphics environment that would allow frequent changes and customization. Thus, we turned to a final environment custom-made for influence-diagram-based expert systems. The KNET system, also developed by Mr. Chavez, separates the inferencing capabilities and graphical manipulation of the knowledge base into MPW Object Pascal from the textual part of the knowledge base and the the evidence collection in HyperCard. This system runs on the Macintosh II with 4 MB of RAM.

The program code is now entirely independent of the knowledge required for reading papers. REFEREE has a new interface that is intuitive and consistent. There is an innovative consultation mode in which questions are presented in free-format menus. The dialogues are mixed-initiative and of mixed levels, allowing the user such options as requesting more detailed questions or cutting off apparently fruitless lines of questioning. With the new REFEREE prototype, the user interacts with the machine using a mouse-pointing device Finally, the screen enables the user to orient himself at all times, obviating the need for special commands to help the user "navigate" through the knowledge base. Our expert recently provided the best indication of the usability of this new system. After only a brief introduction to the new machine and interface, he was able - for the first time - to run an entire consultation by himself.

Current Status:

At this point, REFEREE is a prototype that enables the clinician to read clinical trials more critically. A number of computational issues remain, such as the optimal representation of Dr. Brown's knowledge in our current formalism, and the decision-theoretic extensions. Furthermore, REFEREE represents only the first step in a larger research plan, the automation of knowledge acquisition (see section on Research Plans, below). Current work in the restricted domain of clinical trials will, we hope, illustrate general principles in the design of decision makers that gather expertise from written text and multiple knowledge sources.

D. Relevant Publications

1) Haggerty, J.: *REFEREE and RULECRITIC: Two prototypes for assessing the quality of a medical paper*. REPORT KSL-84-49. Master's Thesis, Stanford University, May 1984.

2) *Chavez, R. Martin and Cooper, G. F.: *KNET: Integration Hypermedia and Normative Bayesian Modeling*. Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence, University of Minnesota, Minneapolis, Minnesota, Aug 19-21, 49-54, 1988.

3) *Lehmann, H. *Knowledge Acquisition for Probabilistic Expert Systems.* Proceedings of the Twelfth Symposium on Computer Applications in Medical Care, Washington, D.C., Nov 6-9, 73-77, 1988.

4) *Lehmann, H. *A Decision-Theoretic Model for Using Scientific Data* Submitted to the Fifth Uncertainty Workshop in Artificial Intelligence, 1989.

E. Funding Support

REFEREE currently receives only a small amount of funding. Most of the research is performed in time contributed by the researchers to this project.

Title: Knowledge-Based Systems Research
PI: Edward A. Feigenbaum
Agency: Defense Advanced Projects Research Agency
Grant identification number: N00039-86-0033
Total award period and amount: 10/1/85 - 9/30/88 $4,130,230 (direct and indirect)
Current award period and amount: 10/1/87 - 9/30/88 $1,467,300 (direct and indirect)
REFEREE component is $27,706, or 1.9 % of grant total.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations

Dr. Brown and Dr. Feldman of the Stanford University School of Medicine are actively involved in the REFEREE project and are the primary domain experts and critics for this project.

C. Critique of Resource Management

The SUMEX computer resource and Lisp workstations have been very important for the work to date, and the SUMEX staff has continued to be very cooperative with the REFEREE project.

III. RESEARCH PLANS

A. Goals & Plans

The overall objective of the REFEREE project is to use recent Artificial Intelligence techniques to build a system that helps the informed but statistically non-expert reader to evaluate critically the medical literature on randomized controlled trials (RCT's). This system will contain and be able to apply dynamically the detailed specialized knowledge of Dr. Byron W. Brown, a biostatistician expert in the design and evaluation of randomized controlled trials. We have divided our overall objective into two goals:

- Goal 1 is the construction of an expert system to help readers (e.g., medical students, medical researchers, clinicians, journal editors, or editorial assistants) assess the credibility of a *single* conclusion drawn from a *single* journal report of a randomized controlled trial. We have already made substantial progress toward this goal with the development of the prototype REFEREE system.
- Goal 2 is the expansion of REFEREE to an expert system that can be used by a similar range of readers to facilitate the evaluation of *multiple* reports based on randomized controlled trials. This expanded system, to be known as the REVIEWER, will thus perform meta-analysis.

The task of extending and refining the prototype REFEREE system in order to achieve these goals can be characterized in terms of three dimensions:

- Making the system more accessible to a variety of people by improving the user interface, validating the system's performance with different types of users, and providing an explanatory capability
- Expanding the knowledge base by continuing the knowledge acquisition process to cover additional types of RCT's
- Improving the inference engine to ensure consistency of the knowledge base and to focus the consultation process on questions relevant to the situation and the individual user.

The specific steps that are planned for the enhancement of the REFEREE system include the following:

- Critique individual clinical trials according to the methodological quality of the trial;
- Measure the efficacy of treatment as demonstrated in a randomized control trial;
- Compare and contrast the credibility and efficacy of treatment reported by multiple journal articles; and
- Combine the *qualitative* techniques of heuristic reasoning and the *quantitative* methods of statistical meta-analysis to extract a consensus opinion from multiple knowledge sources.

In addition, plans for Goal 2, the REVIEWER system to analyze multiple RCT's and form a consensus judgment, include:

- Complete a review of the available literature on meta-analysis and augment the REFEREE prototype to produce estimators for meta-analysis and incorporate expert knowledge on the appropriateness of these methods.
- Add explicit and heuristic knowledge needed for the calculation of robust, non-parametric estimators of effect size.
- Construct a prototype of a system that builds categorical models in the domain of Bayesian meta-analysis, to perform autonomous investigations in the domain of statistical model-building. The REVIEWER will utilize expert knowledge in biostatistics to guide its search for meaningful models.
- Package the REVIEWER in a form suitable for use by physicians and their assistants.
- Verify the expertise of the REVIEWER system on a suite of papers drawn from clinical trials, similar to the validation of REFEREE above.

B. Justification for Continued SUMEX Use

The local area network maintained by the SUMEX staff is essential to the effective development and use of the REFEREE system on Lisp workstations. The connections to local and national computer networks such as ARPANET are important for sharing ideas and results with other medical researchers.

C. Need for other computing resources

REFEREE is currently implemented on the Macintosh II personal computers. We anticipate the need for at least two of these machines for transporting our system and developing new modes of interaction with both naive and experienced users.

IV.D. Pilot AIM Projects

Following is a description of the informal pilot projects currently using the AIM portion of the SUMEX-AIM resource, pending funding, full review, and authorization.

IV.D.1. The Pathfinder Project

Bharat Nathwani, M.D
Department of Pathology
University of Southern California

Lawrence M. Fagan, M.D., Ph.D
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

Our project addresses difficulties in the diagnosis of lymph node pathology. Several studies from cooperative oncology groups have documented that, while experts show agreement with one another, the diagnosis made by practicing pathologists may have to be changed by expert hematopathologists in as many as 50% of the cases. Precise diagnoses are crucial for the determination of optimal treatment. To make the knowledge and diagnostic reasoning capabilities of experts available to the practicing pathologist, we have been exploring issue of representation and inference with expert pathology knowledge. A computer-based diagnostic program called Pathfinder has been developed that is centered on the implementation of principles of probability and decision theory. The project is a collaborative effort of the University of Southern California and the Stanford University Medical Computer Science Group. The most recent version of the program provides diagnostic advice on over 70 common benign and malignant diseases of the lymph node based on over 100 histologic features. The design of the program, with special regard to the hypothetico-deductive reasoning architecture of the Pathfinder system, was influenced by the hypothetico-deductive architecture of the INTERNIST-1/CADUCEUS program developed on the SUMEX resource.

Pathfinder computer-science research is focused on the exploration and extension of formal techniques for decision making under uncertainty Research foci have included (1) the assessment and representation of important probabilistic dependencies among morphologic features and diseases, (2) reasoning about the costs and benefits of alternative information acquisition strategies, (3) the acquisition and use of expert knowledge bases from multiple experts, (4) the customization of the system's reasoning and explanation behaviors to reflect the expertise of the user, and, (5) controlling the naturalness of complex formal reasoning techniques.

Toward the pragmatic goal of constructing a useful pathology teaching and decision-support system, Pathfinder investigators have sought to apply intelligent computation to substantially increase the quantity and quality of pathology knowledge available to pathologists. Important areas of this knowledge integration task involve ongoing research on the crisp definition important morphologic features and feature severities, the synthesis of information from multiple experts, and the translation among multiple pathology classification schemes.

A group of expert pathologists from several centers in the U.S. have showed interest in the program and helped to provide the structure of the knowledge base for the Pathfinder system.

B. Medical Relevance and Collaboration

One of the most difficult areas in surgical pathology is the microscopic interpretation of lymph node biopsies. Most pathologists have difficulty in accurately classifying lymphomas. As mentioned above, several cooperative oncology group studies have documented that while experts show agreement with one another, the diagnosis rendered by a "local" pathologist may have to be changed by expert lymph node pathologists (expert hematopathologists) in as many as 50% of the cases.

The National Cancer Institute recognized this problem in 1968 and created the Lymphoma Task Force which is now identified as the Repository Center and the Pathology Panel for Lymphoma Clinical Studies. The main function of this expert panel of pathologists is to confirm the diagnosis of the "local" pathologists and to ensure that the pathologic diagnosis is made uniform from one center to another so that the comparative results of clinical therapeutic trials on lymphoma patients are valid. An expert panel approach is only a partial answer to this problem. The panel is useful in only a small percentage (3%) of cases; the Pathology Panel annually reviews only 1,000 cases whereas more than 30,000 new cases of lymphomas are reported each year. A panel approach to diagnosis is not practical and lymph node pathology cannot be routinely practiced in this manner.

We believe that practicing pathologists do not see enough case material to maintain a high level of diagnostic accuracy. The disparity between the experience of expert hematopathology teams and those in community hospitals is striking. An experienced hematopathology team may review thousands of cases per year. In contrast, in a community hospital, an average of only ten new cases of malignant lymphomas are diagnosed each year. Even in a university hospital, only approximately 100 new patients are diagnosed every year.

Because of the limited numbers of cases seen, pathologists may not be conversant with the differential diagnoses consistent with each of the histologic features of the lymph node; they may lack familiarity with the complete spectrum of the histologic findings associated with a wide range of diseases. In addition, pathologists may be unable to fully comprehend the conflicting concepts and terminology of the different classifications of non-Hodgkin's lymphomas, and may not be cognizant of the significance of the immunologic, cell kinetic, cytogenetic, and immunogenetic data associated with each of the subtypes of the non-Hodgkin's lymphomas.

In order to promote the accuracy of the knowledge base development we will have participants for multiple institutions collaborating on the project. Dr.

Nathwani will be joined by experts from Stanford (Dr Dorfman), St. Jude's Children's Research Center -- Memphis (Dr Berard) and City of Hope (Dr. Burke).

C. Highlights of Research Progress

C.1 Overview

Pathfinder research, has centered on the development of tractable methods for the acquisition, representation, and inference with probabilistic knowledge in pathology. Two M.D./Ph.D (Stanford Medical Information Science Program) students, David Heckerman and Eric Horvitz, designed and implemented the program and have played a central role in the direction of research on the project. In the past five years, the Pathfinder team has worked to (1) build a large consensus knowledge base of probabilistic inference, (2) to refine techniques of hypothetico-deductive reasoning, (3) to develop techniques for modulating the complexity of formal inference to enhance the clarity of reasoning and explanation, and (4) to begin formal evaluation of the performance of the system. Some of the Pathfinder research has stimulated other expert-systems research efforts centering on the re-examination of the construction of systems grounded in the principles of probability and decision theory.

C.2 History of Pathfinder System Implementation

Since the project's inception in September, 1983, we have constructed several versions of Pathfinder. The first several versions of the program were rule-based systems like MYCIN and ONCOCIN which were developed earlier by the Stanford group. These systems were implemented in the MRS logic theorem-proving language. We discovered early-on, however, that the large number of overlapping features in diseases of the lymph node would make a rule-based system cumbersome to implement. We next considered the construction of a hybrid system, consisting of a rule-based algorithm that would pass control to an INTERNIST-1-like scoring algorithm if it could not confirm the existence of classical sets of features. Later we, applied formal probabilistic representation and reasoning methods in a hypothetico-deductive reasoning framework. The original version of Pathfinder is written in the computer language MacLisp and runs on the SUMEX DEC-2060. This was transferred to Portable Standard Lisp (PSL) on the DEC-2060, and later transferred to PSL on the HP 9836 workstations. Two years ago, the Pathfinder team reimplemented the program in MPW Object Pascal on the Macintosh II. Much of the recent testing and refinement of the knowledge base has been carried out within the Macintosh II environment.

C.3 Pathfinder Knowledge Base

Initial versions of the Pathfinder knowledge base was constructed by Dr. Nathwani. During the early part of 1984, we organized two meetings of the entire team, including the pathology experts, to define the selection of diseases to be included in the system, and the choice of features to be used in the scoring process. During the last three years, we have focused on methodologies for more accurately representing expert knowledge about the uncertain relationships between features and diseases in lymph-node pathology. Early versions of the Pathfinder knowledge base assumed independence between features used in diagnosis. However, knowledge-engineering sessions with the PI, who served as the chief hematopathology expert on the Pathfinder team, identified important probabilistic dependencies among features used in lymph node pathology. We have pursued the representation of the uncertain causal and associational relationships among features and diseases in lymph node pathology. We have found that attempting to move beyond the assumptions of conditional independence does not necessarily lead to an exponential growth in the tasks of knowledge acquisition, representation, and inference.

We have addressed the problem of probabilistic dependencies with a promising representation, developed in the decision science community, called belief networks. Although belief networks have been used as an alternative to decision trees for performing single analyses, there has been relatively little experience with the use of this representation in expert systems development. We pursued the use of belief networks because of the representation's soundness and expressiveness. With belief networks, probabilities are used to quantitate the beliefs about qualitative dependencies asserted by the expert. We found the belief network to be an intuitive and practical representation for building a large knowledge base. We have worked to enrich the basic belief network representation by developing a new language and associated operators for describing new types of conditional independence among findings. We found that a graphical knowledge-acquisition technique, called similarity-networks, could facilitate the knowledge acquisition process for building large, probability-based knowledge bases.

C.4 Simplification of Probabilistic Reasoning and Explanation

We have also focused on the problem of making complex information-theoretic inference understandable and explainable. We found that straightforward applications of decision-theoretic inference could lead to computer problem-solving behavior viewed as confusing or counterintuitive to users. Early, less-flexible versions of Pathfinder worked solely on the finest distinctions available in the system's representation. We found that users tended to work at higher levels of abstraction than did our straightforward decision-theoretic approach. Users also preferred to make specific transitions from one subproblem to another.

Knowledge acquisition with several pathologists unearthed alternative problem-solving control hierarchies that seemed to be used to segment a single complex diagnostic reasoning task (from the perspective of the decision-theoretic system) into a set of tasks at increasingly detailed levels of abstraction. These human-oriented abstraction strategies are useful for allowing a pathologist to reason about groups of similar diseases rather than consider each disease as a separate entity. We have worked to acquire and apply alternative control strategies from trainees and experts. We worked to enhance the Pathfinder system to enable a user to probe a differential diagnosis from alternative perspectives. The current system allows a user to dynamically select alternative strategies for grouping the current differential list.

C.5 Evaluation of Pathfinder Performance

We applied a heuristic and decision-theoretic metric to perform a comparative analysis of the importance of enriching a conditional independence model with dependency knowledge. The study compared the performance of the system with that of the domain expert. In the evaluation study, a community pathologist used the Pathfinder system to analyze a set of difficult cases. Fifty-three cases were were selected in sequence from a large library of referrals. As each case was entered into the system, probability distributions over disease hypotheses or differentials were generated by Pathfinder. In the next phase of the evaluation, the diagnostic accuracy of the distribution produced by Pathfinder was gauged by assigning it a score based on two metrics. We applied a heuristic scoring approach and a formal decision-theoretic approach. We found the two approaches to be complementary in their ability to identify components of system performance. The work showed a close correspondence between the behavior of the system and expert decision making.

C.6 SUMEX Usage

Although the SUMEX-AIM Resource was central in the initiation of the Pathfinder project, and in the prototyping of the early Pathfinder expert systems, the system not been used directly for development over the last three years. The resource has been used during this time for electronic mail and file archiving. Nevertheless, this SUMEX-AIM service has played a central role in communication among the participants on the Pathfinder project, especially for facilitating communication between the Stanford and USC Pathfinder research teams.

D. Publications Since January 1984

1) Horvitz, E. J., Heckerman, D. E., Nathwani, B. N. and Fagan, L. M.: "Diagnostic Strategies in the Hypothesis-directed Pathfinder System, Node Pathology." HPP Memo 84-13. Proceedings of the First Conference on Artificial Intelligence Applications, Denver, Colorado, Dec., 1984.

2) Heckerman, D. E., and Horvitz, E. J., "The Myth of Modularity in Rule-based Systems," in Uncertainty in Artificial Intelligence, Vol 2, J. Lemmer, L. Kanal, ed., North Holland, New York, 1987.

3) Horvitz, E. J., Heckerman, D. E., Nathwani, B. N. and Fagan, L. M.: "The Use of a Heuristic Problem-solving Hierarchy to Facilitate the Explanation of Hypothesis-directed Reasoning." KSL Memo 86-2 Proceedings of MedInfo, Washington D.C., October, 1986.

4) Horvitz, E. J., "Toward a Science of Expert Systems," Invited Paper, Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface, American Statistical Association, March, 1986, pgs 45-52.

5) Heckerman, D. E., "An Axiomatic Framework for Belief Updates," in Uncertainty in Artificial Intelligence, Vol. 2, J. Lemmer, L. Kanal, ed., North Holland, New York, 1987.

6) Heckerman, D. E., and Horvitz, E. J., "The Myth of Modularity in Rule-based Systems," in Uncertainty in Artificial Intelligence, Vol 2, J. Lemmer, L. Kanal, ed., North Holland, New York, 1987.

7) Heckerman, D. E., and Horvitz, E. J., "On the expressiveness of rule-based systems for reasoning under uncertainty," Proceedings of the National Conference on Artificial Intelligence, Seattle, Washington, July, 1987.

8) Horvitz, E. J., Heckerman, D. E., Langlotz, C. P., "A framework for comparing alternative formalisms for plausible reasoning," Proceedings of the AAAI," August, 1986, Morgan Kaufman, Los Altos, CA, 1986.

9) Horvitz, E.J., Breese, J.S., Henrion, M., Decision Theory in Expert Systems and Artificial Intelligence, International Journal of Approximate Reasoning, Elsevier, N.Y. July, 1988.

10) Heckerman, D.E., An Evaluation of Three Scoring Schemes, Proceedings of the 4th AAAI Workshop on Uncertainty in Artificial Intelligence, Minneapolis, MN., (to appear August 1988).

11) Horvitz, E.J., A Multiattribute Utility Approach to Inference Understandability and Explanation, Tech. Report, KSL-28-87, Knowledge Systems Laboratory, Stanford, California, March, 1987.

12) Horvitz, E.J., "Reasoning About Beliefs and Actions Under Computational Resource Limitations," AAAI Workshop on Uncertainty in Artificial Intelligence, Seattle, Washington, 1987.

13) Horvitz, E.J., "Problem-solving Design: Reasoning About Computational Value, Resources, and Tradeoffs," Proceedings of the NASA Artificial Intelligence Forum. Palo Alto, California, November,1987.

14) Nathwani, B.N., Horvitz, E.J., Heckerman, D.E., Lincoln, T., "Expert Systems and Interactive Videodiscs in Diagnostic Pathology: Augmenting the Multidisciplinary Approach," Human Pathology. In press.

15) Heckerman, E.J. Horvitz, B.N. Nathwani, "Toward Effective Normative Decision Systems: Update on the Pathfinder Project", Technical Report KSL-89-25, March,1989. Knowledge Systems Laboratory, Stanford, CA; submitted to SCAMC-1989.

16) Horvitz, D.E. Heckerman, K. Ng, B.N. Nathwani, "Heuristic Abstraction in the Decision-Theoretic Pathfinder System," Technical Report KSL-89-24, March,1989. Knowledge Systems Laboratory, Stanford, CA; submitted to SCAMC-1989.

E. Funding Support

Research Grant submitted to National Institutes of Health Grant
Title: "Computer-aided Diagnosis of Malignant Lymph Node Diseases"
Principal Investigator: Bharat Nathwani
Funding for three years from the National Library of Medicine
1 RO1 LM 04529 $766,053 (direct and indirect)

Professional Staff Association, Los Angeles County Hospital, $10,000

University of Southern California, Comprehensive Cancer Center, $30,000

Project Socrates, Univ. of Southern Calif., Gift from IBM of IBM PC/XT.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Program Dissemination via SUMEX

Because our team of experts are in different parts of the country and the computer scientists are not located at the USC, we have made use of SUMEX for communication, demonstration of programs, and remote modification of the knowledge base.

B. Sharing and Interaction with Other SUMEX-AIM Projects

We have been in touch with other sites interested in Pathfinder research. As an example, the SUMEX pilot project, RXDX, designed to assist in the diagnosis of psychiatric disorders, is currently using a version of the Pathfinder program on the DEC-2060 for the development of early prototypes of future systems.

C. Critique of Resource Management

The SUMEX resource has provided an excellent basis for the development of a pilot project. The availability of a pre-existing facility with appropriate computer languages, communication facilities (especially the TELENET network), and document preparation facilities allowed us to make good progress in a short period of time. The management has been very useful in assisting with our needs during the start of this project.

III. RESEARCH PLANS

A. Project Goals and Plans

The current Pathfinder research grant will come to an end in Fall, 1989. The Pathfinder team is currently seeking a new grant to support a formal multicenter clinical trial to ascertain the efficacy of the use of system based on the Pathfinder knowledge base and inference techniques. We plan to carry out a randomized trial of the use of the system with general pathologists. In addition to the statistical analysis of the efficacy of the system for providing assistance with the diagnostic subproblems of feature identification and integration, the group plans to study pathologists' attitudes on the use of computer-based decision support systems in the clinical environment.

B. Requirements for Continued SUMEX Use

We are currently dependent on the SUMEX computer for file storage and archival, and for communication. While the switch to workstations has lessened our requirements for computer time for the development of the algorithms, we will continue to need the SUMEX facility for the interaction with each of the research locations specified in our NIH proposal. An early version of the Pathfinder systems is stored on the SUMEX mainframe for use by non-Stanford users.

C. Requirements for Additional Computing Resources

Most of our computing resources will be met by the use of the Macintosh II workstations. However, we will continue to need additional file space on the SUMEX system for our continuing development and clinical trials work. We will also continue to require access to SUMEX for communication purposes, access to other programs, and for file storage and archiving.

Appendix A: Knowledge Systems Laboratory Brochure

ARTIFICIAL INTELLIGENCE RESEARCH IN THE
KNOWLEDGE SYSTEMS LABORATORY

Stanford University
Department of Computer Science
Department of Medicine
March 1989

Introduction

The Knowledge Systems Laboratory (KSL) is an artificial intelligence (AI) research laboratory of approximately 100 people—faculty, staff, and students—within the Departments of Computer Science and Medicine at Stanford University. KSL is the name for the interdisciplinary AI research community that has evolved over the past two decades. Begun as the DENDRAL Project in 1965 and known as the Heuristic Programming Project from 1972 to 1984, the new organization reflects the diversity of the research now under way. The KSL is a modular laboratory, consisting of three collaborating yet distinct groups with different research themes:

- **The Heuristic Programming Project (HPP)**, Professor Edward A. Feigenbaum, scientific director (Department of Computer Science)—large, multi-use knowledge bases, blackboard systems, concurrent system architectures for AI, automated software design, expert systems for science and engineering. Executive director: Robert Engelmore. Research scientists: Harold Brown, Bruce Delagi, Barbara Hayes-Roth, Yumi Iwasaki, Tom Gruber, Richard Keller, Hirotoshi Maegawa, Penny Nii, and Kazuo Tanaka.
- **The Medical Computer Science (MCS) Group**, Associate Professor Edward H. Shortliffe, scientific director (Department of Medicine with courtesy appointment in Computer Science)—fundamental research and advanced biomedical applications in the area of AI and decision sciences; includes the Medical Information Sciences (MIS) program. Assistant Professor: Mark A. Musen. Associate Director: Lawrence M. Fagan. Research scientist: Gregory F. Cooper.
- **The Symbolic Systems Resources Group (SSRG)**, Thomas C. Rindfleisch, scientific director (joint appointment Departments of Computer Science and Medicine)—development of distributed computing environments for AI research and operation of KSL computing resources, including the SUMEX-AIM facility. SSRG Group Leaders: Richard Acuff, Christopher Lane, Nicholas Veizades, and William J. Yeager.

The KSL is guided by an Executive Committee consisting of the three sublaboratory directors and administrative managers. Tom Rindfleisch serves as overall KSL director (see Figure 1).

This brochure summarizes the goals and methodology of the KSL, its research and academic programs, its achievements, and the research environment of the laboratory.

Basic Research Goals and Methodology

Throughout a 24-year history, the KSL and its predecessors, DENDRAL and HPP, have concentrated on research in expert systems—that is, systems using symbolic reasoning and problem-solving processes that are based on extensive domain-specific knowledge. The KSL's approach has been to focus on applications that are themselves significant real-world problems (in domains such as science, medicine, engineering, and education), and that also expose key, underlying AI research issues. For the KSL, AI is largely an empirical science. Research problems are explored, not by examining strictly theoretical questions, but by designing, building, and experimenting with programs that serve to test underlying theories.

The basic research issues at the core of the KSL's interdisciplinary approach center on the computer representation and use of large amounts of domain-specific knowledge, both factual and heuristic (or judgmental). These questions have guided our work since the 1960's and are now of central importance in all of AI research:

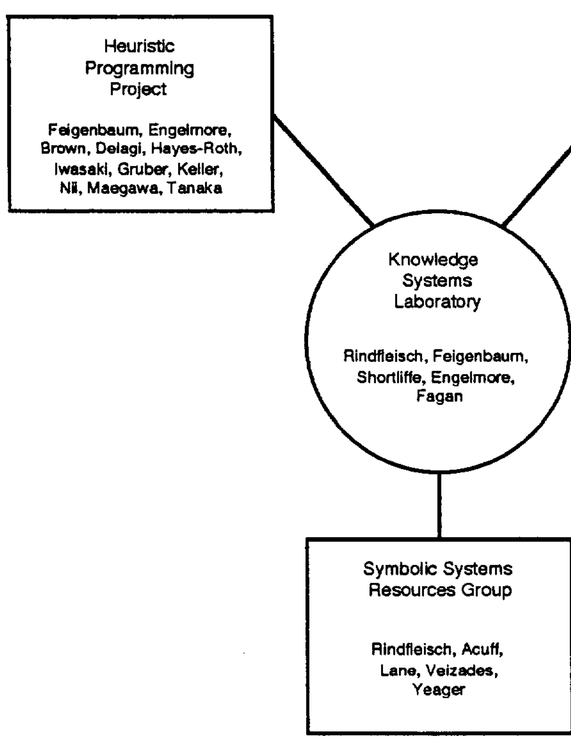

Figure 1 — Knowledge Systems Laboratory Organization

1. **Knowledge representation**. How can the knowledge necessary for complex problem solving be represented for its most effective use in automatic inference processes? Often, the knowledge obtained from experts is heuristic knowledge, gained from many years of experience. How can this knowledge, with its inherent vagueness and uncertainty, be represented and applied? How can knowledge be represented so that it can be used for many problem solving purposes? Can knowledge be abstracted for use in multiple ways?

2. **Knowledge acquisition**. How is knowledge acquired most efficiently—whether from human experts, from observed data, from experience, or by discovery? How can a program discover inconsistency and incompleteness in its knowledge base? How can knowledge be added without perturbing the established knowledge base unnecessarily?

3. **Use of knowledge.** By what inference methods can many sources of knowledge of diverse types be made to contribute jointly and efficiently toward solutions? How can knowledge be applied at the appropriate time and at the appropriate level of detail? How can existing knowledge be transformed so it is suitable for use by a specific application task?

4. **Explanation and tutoring**. How can the knowledge base and the line of reasoning used in solving a particular problem be explained to users? What constitutes a sufficient or an acceptable explanation for different classes of users?

5. **System tools and architectures**. What kinds of software tools and system architectures can be constructed to make it easier to implement expert programs with greater complexity and higher performance? What kinds of systems can serve as vehicles for the cumulation of knowledge of the field for the researchers? What architectural properties enable a system to function in real-time task environments?

Current Research Projects

The following is a summary of projects now under way within the three KSL research groups and gives the major goals of each project and lists the personnel (staff and Ph.D. candidates) directly involved. More complete information on individual projects can be obtained from the person indicated as the project contact. Inquiries should be addressed in care of:

Knowledge Systems Laboratory
Department of Computer Science
Stanford University
701 Welch Road, Building C
415-723-3444

The Heuristic Programming Project

- **Advanced Architectures Project**—Design a new generation of computer hardware architectures and problem solving frameworks to exploit concurrency in knowledge-based signal understanding systems.
  Personnel: Edward A. Feigenbaum (contact), Nelleke Aiello, Harold Brown, Bruce Delagi (DEC), Robert Engelmore, Hirotoshi Maegawa (Sony), Penny Nii, Sayuri Nishimura, James Rice, Nakul Saraiya.
- **Blackboard Architecture for Adaptive Intelligent Systems**—Design and develop a software architecture for systems that must reason about and interact with dynamic external entities in real time. Includes the Guardian project to develop a prototype system for real-time monitoring of surgical intensive care patients (see related VENTPLAN project under the Medical Computer Science Group).
  Personnel: Barbara Hayes-Roth (contact), Richard Washington, Rattikorn Hewett, Adnan Darwiche, Michael Wolverton, Andrew Gans, Anthony Confrey, Luc Boureau, Anne Collinot, Iris Tommelein, Edward Chang, James Rice, Adam Seiver (Palo Alto VAMC).
- **Large, Multi-use Knowledge Base (LMKB) Project**—Develop an expert systems architecture capable of supporting multiple application tasks involving reasoning about engineered devices (e.g., device monitoring, diagnosis, redesign, assembly, instruction), using a large, common knowledge base of science and engineering principles underlying device design and operation.
  Personnel: Edward Feigenbaum (contact), Richard Keller, Robert Engelmore, Yumi Iwasaki, Kazuo Tanaka (NTT), Tom Gruber.
- **Automated Software Design and Redesign**—Assist software designers in designing new systems via intelligent selection, modification, and construction from a library knowledge base of existing software modules.
  Personnel: Penny Nii (contact), Cordell Green (Kestrel Institute), Nelleke Aiello, Raul Duran, Liam Peyton.

The Medical Computer Science Group

- **ONCOCIN**—Develop knowledge-based systems for the administration of complex medical treatment protocols such as those encountered in cancer chemotherapy.
  Personnel: Ted Shortliffe (contact), Charlotte Jacobs (Oncology), Larry Fagan, David Combs, Robert Carlson, Christopher Lane, Rick Lenon, Mark Musen, Janice Rohn, Samson Tu, Cliff Wulfman, Andrew Zelenetz.
- **OPAL/PROTÉGÉ**—Develop graphics-based knowledge acquisition tools for clinical trials. OPAL developed out of the ONCOCIN project to provide a method for specifying cancer treatment experiments. The PROTÉGÉ program is capable of creating OPAL-like knowledge acquisition tools for various areas of medicine.
  Personnel: Mark Musen (contact), David Combs.

- **Speech Input to Expert Systems**—Develop multi-modal interface to expert systems, concentrating on a connected speech input device. Primary application will be extension to the ONCOCIN graphical interface.
  Personnel: Larry Fagan (contact), Bonnie Webber (University of Pennsylvania), Ted Shortliffe, Ed Feigenbaum (HPP), Ellen Isaacs (Psycholinguistics), Monica Rua, Clifford Wulfman, Christopher Lane, Janice Rohn.
- **Physician's Workstation**—Develop advanced integrated workstation suitable for providing decision support functions to clinicians in both inpatient and outpatient settings.
  Personnel: Ted Shortliffe (contact), Tom Rindfleisch, Clifford Wulfman.
- **Qualitative and Quantitative Computation (VENTPLAN)**—Develop methods to combine qualitative and quantitative processing techniques in order to interpret and react to data gathered in time-varying application areas. The VENTPLAN system interprets data from the Intensive Care Unit, and suggests settings for mechanical ventilators (see related Guardian project in HPP).
  Personnel: Larry Fagan (contact), Adam Seiver (Palo Alto Veterans Hospital), Lewis Sheiner (University of California, San Francisco), Ingo Beinlich, Brad Farr, Jeanette Polaschek, John Reed, Geoff Rutledge, George Thomsen, Samson Tu.
- **Decision-Theoretic Expert Systems**—Develop pragmatic methods of knowledge acquisition, inference, and explanation for medical expert systems based on decision theory.
  Personnel: Greg Cooper (contact), Ted Shortliffe, Martin Chavez, David Heckerman, Edward Herskovits, Eric Horvitz, Harold Lehmann, Richard Lin, Blackford Middleton, Mike Shwe, Jaap Suermondt.

The Symbolic Systems Resources Group (SSRG)

- **SUMEX-AIM Resource**—Develop and operate a national computing resource for biomedical applications of artificial intelligence in medicine and for basic research in AI at KSL.
  Personnel: Tom Rindfleisch (contact), Rich Acuff, Frank Gilmurray, Christopher Lane, Christopher Schmidt, Andrew Sweer, Bob Tucker, Nicholas Veizades, Bill Yeager.
- **AI Workstation and Network Systems**—Develop network-based computing environments for AI research on workstations including remote graphics and distributed computing.
  Personnel: SSRG staff

Students and Special Degree Programs

Graduate students are an essential part of the research productivity of the KSL. Currently 30 students are working with our projects centered in Computer Science and another 24 students are working with the MCS/MIS programs in Medicine. Because of the highly interdisciplinary and experimental nature of KSL research, a special degree program, the *Medical Information Sciences (MIS)* program, was approved by Stanford University in 1982. It offers instruction and research opportunities leading to the M.S. or Ph.D. degree in medical information sciences. The program, directed by Ted Shortliffe and co-directed by Larry Fagan, is formally administered by the School of Medicine, but the curriculum and degree requirements are coordinated with the Dean of Graduate Studies and the Graduate Studies Committee of the University. The program reflects our local interest in the interconnections between computer science, artificial intelligence, and medical problems. Emphasis is placed on providing trainees with a broad conceptual overview of the field and with an ability to create new theoretical and practical innovations of clinical relevance. Of the 24 MIS students currently, 17 are working toward Ph.D. degrees, and 7 are working toward M.S. degrees.

Academic and Research Achievements

The primary products of our research are scientific publications on the basic research issues that motivate our work, computer software in the form of the expert systems and AI architectures we develop, and the students we graduate who continue AI research in other academic and industrial laboratories.

The KSL has averaged publishing more than 45 research papers per year in the AI literature, including journal articles, theses, proceedings articles, and working papers.[1] In addition, many talks and invited lectures are given annually. In the past few years, 12 major books have been published by KSL faculty, staff, and former students, and several more are in progress. Those recently published include:

- *Automated Generation of Model-Based Knowledge-Acquisition Tools*, Musen, Pitman, 1989.
- *Blackboard Systems*, Engelmore and Morgan, eds., Addison-Wesley, 1988
- *The Rise of the Expert Company: How Visionary Companies are Using Artificial Intelligence to Achieve Higher Productivity and Profits*, Feigenbaum, McCorduck and Nii, Times Books, 1988.

[1] Copies of individual KSL publications may be obtained through the Stanford Department of Computer Science Publications Office. The full collection of KSL reports has been published in microfiche by COMTEX Scientific Corporation.

- *A Computational Model of Reasoning from the Clinical Literature*, Rennels, Lecture Notes in Medical Informatics, Volume 32, Springer-Verlag, 1987.
- *Heuristic Reasoning about Uncertainty: An AI Approach*, Cohen, Pitman, 1985.
- *Readings in Medical Artificial Intelligence: The First Decade*, Clancey and Shortliffe, Addison-Wesley, 1984.
- *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Buchanan and Shortliffe, Addison-Wesley, 1984.
- *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, Feigenbaum and McCorduck, Addison-Wesley, 1983.
- *Building Expert Systems*, F. Hayes-Roth, Waterman, and Lenat, eds., Addison-Wesley, 1983.
- *System Aids in Constructing Consultation Programs: EMYCIN*, van Melle, UMI Research Press, 1982.
- *Knowledge-Based Systems in Artificial Intelligence: AM and TEIRESIAS*, Davis and Lenat, McGraw-Hill, 1982.
- *The Handbook of Artificial Intelligence*, Volume I, Barr and Feigenbaum, eds., 1981; Volume II, Barr and Feigenbaum, eds., 1982; Volume III, Cohen and Feigenbaum, eds., 1982; Kaufmann.
- *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*, Lindsay, Buchanan, Feigenbaum, and Lederberg, McGraw-Hill, 1980.

Our laboratory has pioneered in the development and application of AI methods to produce high-performance knowledge-based programs. Programs have been developed in such diverse fields as analytical chemistry (DENDRAL), infectious disease diagnosis and treatment (MYCIN), cancer chemotherapy management (ONCOCIN), pulmonary function evaluation (PUFF), VLSI design (KBVLSI/PALLADIO), molecular biology (MOLGEN), parallel machine architecture simulation (CARE), and parallel problem solving (POLIGON). Some of our systems and tools (e.g., UNITS, EMYCIN, and AGE) have been adapted for commercial development and use in the AI industry.

Following our lead in work on biomedical applications of AI and the development of the SUMEX-AIM computing resource, a nationally recognized community of academic projects on AI in medicine has grown up.

KSL faculty, staff, and students have been recognized internationally for the quality of their work and for their continuing contributions to the field. KSL members participate extensively in professional organizations, government advisory committees, and journal editorial boards. They have held